

A Study of Item Response and Anchor Bias in Economics Objective Tests among Senior Secondary School Students in Osun State

Ajeigbe Taiwo Oluwafemi

Department of Educational Foundations and Counselling, Faculty of Education,
Obafemi Awolowo University, Ile-Ife, Nigeria
taiaje215@yahoo.com

Afolabi Eytayo Rufus Ifedayo

Department of Educational Foundations and Counselling, Faculty of Education,
Obafemi Awolowo University, Ile-Ife, Nigeria
eriafolabi@oauife.edu.ng

Doi:10.5901/ajis.2013.v2n3p173

Abstract

The study investigated the extent to which test option keys exhibit middle bias, the influence of primacy, recency, and middle attractions on item difficulty and discrimination in multiple-choice items with a view to improving the quality of objective test. The population for the study consisted of Senior Secondary School three (SSSIII) students in Osun State. The sample consisted of 620 SSSIII Economics Students randomly selected from 20 randomly selected schools in five purposively selected local government areas. The research instruments used for the study were a 50-item Economics Multiple-Choice (MC) test adopted from WAEC questions with 4-option length and its adapted equivalent MC test with 5-options length. The results revealed that correct answers were placed in the middle positions in 49.2% of the times in 4-option MC test items as against 64.4% in the 5-option MC test. Also, primacy and recency effects were significant on item difficulty in 4-option MC test ($\chi^2=164.16$; $p<.05$) and the 5-option MC test ($\chi^2=39.69$; $p<.05$). Further, the results showed that middle attraction had significant effect on item difficulty in 5-option MC ($\chi^2=519.50$; $p<.05$) but had a not significant effect in the 4-option MC test ($\chi^2=3.66$; $p>.05$). In addition, middle attraction had significant effect on item discrimination in both test formats ($\chi^2=41.93$; $p<.05$) ($\chi^2=135.63$; $p<.05$). The study concluded that test developers tended to exhibit middle bias in placing correct keys in 5-option MC format than in 4-option MC and that key balancing was more pronounced in 4-option MC test.

Keywords: item response; anchor bias

1. Introduction

An objective test is a type of test in which the score is independent of the subjective influence of the scorer; that is, the score is consistent regardless of the bias or prejudice of the scorer. Essentially, what is 'objective' in objective test is the consistency of the test result, and not the form of the test. In objective tests, there are clear right and wrong answers, as well as fixed marks for each answer. Objective tests represent an efficient way of assessing a large number of students rapidly and within short time frames, particularly when computers are used in marking. Among the various objective test formats, the multiple-choice format is the most widely used in certificate and entrance examinations, standardized tests and in school-based assessments.

The multiple-choice objective test provides the most useful measure of learning outcome in relation to subject content. It can also vary from simple to complex for assessment of different learning outcomes. The typical multiple-choice objective test consists of items that widely sample a subject syllabus or course outline within a well-defined range of cognitive levels. Multiple-choice tests consist of items that require the test taker to select from a set of options. In contrast, constructed response tests consist of items that require the test taker to generate an answer rather than select from a set of options (Bennett, 1993) and encompass a wide variety of response formats ranging from filling in blanks, sentence completion, and short answers to multi-page essays. In multiple-choice items, students must select the correct answer from a given number of possible answers, called foils alternatives or options. The incorrect options are termed distracters, while, the correct option is called the key. Usually, the distracters embody misconceptions, partly correct answers and common errors of fact or reasoning that can turn the attention of unprepared and ill-prepared students away

from the correct answers.

In constructing or writing multiple-choice questions, the test writers usually look for where to hide or place option keys. In an attempt to do this, they may consciously or unconsciously hide the key in a particular position repeatedly while developing the alternatives of multiple-choice test. Sometimes it is a clear case of middle bias. That is placing the option key in a central position as much as up to 3 to 4 times than other positions in the multiple-choice tests. Surprisingly, writers are not aware of this tendency (Attali & Bar-Hillel, 2003). Banks of multiple-choice questions therefore usually exhibit a bias of answers in middle positions. If the option keys are not reassigned to different positions, the resulting option key could be heavily unbalanced. The popular method of dealing with this bias is through the so-called "delicate art of key balancing". Key balancing is not an openly practiced policy.

Mentzer (1982) reported that the distributions of correct answers over positions in a four-option multiple-choice tests revealed that answers are more often placed in a central position. On the other hand, the extent of the biasness between the middle positions and edges positions according to & Attali & Bar-Hillel (2003) was found to be significantly higher than expected by chance (52% in the four-choice tests and 64% in the five-choice tests). The variation recorded around this proportion was relatively small.

The major concern among psychometricians is the order of writing the answers, which depends on how they occur to the writer (i.e. anchoring/anchor bias). The stem and its correct key seem to be a natural point of departure for a multiple – choice item. In this case, the writer may decide to write them down and there after construct distraction should this word, there will be a preponderance of correct answers i.e. more than $1/k$ in the first position. If this trend thus continues, the writer will find it difficult where the last distracter should be placed and this may eventually occupy the last slot. Thereby leading to a preponderance of foil answers i.e. more than $(k-1)/k$ in the last position. Anchoring might lead to a different order, due to the tendency to bracket the correct number (the anchor) by destructors' both large and smaller (Inbar, 2002). Inbar 2002 also pointed it out that the tendency for test writers to exhibit anchor bias is the first position is higher than any other position if the test writers decide to write the correct answer before other distracters.

The numerical answer options are often re-ordered monotonically (Haladyna & Downing, 1989) thus disguising the original order in which they came to mind, the correct answer would be found in central positions more often than $(k-2)/k$. Test takers on the other hand are trying to find out where the option key is located in a set of multiple-choice questions. According to authors like Jerome (1986); Rogers (1991); and (Clark 1956), test takers do show some degree or level of preference for particular option position at the expense of the others. Their findings reveal that the possible position to find the option keys in a multiple-choice test is subject to the level of the difficulty level of the test items. Authors like Aghassi, (1994); and Fagley, (1997) had a contrary opinion that multiple-choice tests are relatively free from position response bias.

Examining the inconsistency of these findings, Jessell and Sullins (1999) argued that selection of option key is not only a function of the understanding of content material in the multiple-choice questions, but rather influenced essentially by the first and last choice that are more outstanding than the middle position in both four and five-option length. The possibility to select either the first or the last option will be higher especially when a test taker fails to go through the alternatives the second time before making a selection.

It was recognised that not only may test takers show favouritism for certain positions, but they might also affect the test writer in his placement of alternatives. Should the same positional factors manifest themselves in both the examinee's pattern of preferences and the test constructor's arrangement of alternatives, the probability of correct responses would be affected. To control this bias, an equal number of correct choices used as anchors (keys) should be randomly scattered among the four and five positions of the multiple-choice tests.

However, regardless of the versatility and popularity of multiple-choice tests, performance on it is influenced by response set, a consistent tendency of test takers to select a particular foil position, to work for speed rather than accuracy, to gamble when in doubt (Afolabi, 1992; Ciardi, 2001; Gall and Spurthein, 1999). Types of response set have been classified to include acquiescence, social desirability, response bias, guessing, and deviant response, among others (Matfessel and Sax, 1958). Response sets are reported to be most apparent when items are ambiguous or when they increase in level of difficulty (Gray, William and Henry 2002).

Response bias is a type of response set that reflects the tendency to select one response position in multiple-choice tests more frequently, not minding the item content of the test. It is also the tendency of test takers to respond in a specific way regardless of test content. Response set is known to influence performance in objective tests and this may be due to testwiseness.

Arising from the above, this study examined the extent to which test option keys exhibit middle bias, primacy and recency effects, as well as effect of middle attraction on item difficulty and item discrimination of four-option and five-

option multiple-choice Economics tests among Senior Secondary School Students in Osun State

2. Statement of the Problem

In objective tests particularly in four-option and five-option multiple-choice tests, it is hard to say preponderance in which test developers favour middle positions or bias to middle positions as the place to hide the option keys (Bar-Hillel and Attali, 2002). The middle bias in tests is a manifestation of the general phenomenon of edge aversion which in spite of some attempts (Shaw, Berger, Brown, & Gallagher, 2000) is still in search of an explanation. The possible position tendencies and item response still remain major areas needed to be investigated. It is not also clear whether students' testwiseness is as a result gender differences, or variation in schools. It is also not known whether the items whose keys are placed in the middle positions will have any influence or effect on the item discrimination of the Multiple-Choice test items than those that are placed in the first or last position in both the four and five option length of multiple-choice tests.

3. Purpose of the Study

The study was conducted to:

- i. determine the extent to which test answer keys exhibit middle bias in four – option and five – option multiple – choice test;
- ii. investigate primacy and recency effects in multiple-choice test on item difficulty and;
- iii. examine the effects of middle attraction on items difficulty and discrimination.

4. Research Question

- i. Do test option keys exhibit middle bias or edge bias in four-option and five-option of the Multiple-Choice test items?

5. Research Hypotheses

- i. There is no significant influence of primacy and recency positions on the item difficulty of the multiple-choice test items.
- ii. There is no significant effect of middle attraction on the item difficulty of the multiple-choice test items.
- iii. There is no significant effect of middle attraction on the discrimination level of the multiple-choice test items.

6. Methodology

The population for the study consisted of all the students who offered Economics at level three in the Senior Secondary Schools in all the 30 Local Government Areas (LGAs) in Osun State. Five Local Government Areas were randomly selected consisting of 104 schools with 11,240 SSSIII Economics students. From the 104 schools, 8 private and 12 public schools were selected using stratified random sampling technique using school ownership as the strata. From each school, intact class was used and thereafter, 31 answer sheets of students that sat for the test were randomly selected before marking. Thus, the total sample size was 620. Two research instruments were used for the study. The first was Economics Multiple-Choice (MC) test adopted from WAEC questions. In this MC test each item has 4-options. The second was an adapted equivalent MC test with 5-options. The internal reliability coefficients of the adapted MC test yielded 0.77.

7. Results

Research Question 1: Do test option keys exhibit middle bias or edge bias in four-option and five-option of the Multiple-Choice test items?

Table 1: Percentage distribution of option keys in WAEC and NECO 4-option length and 5-option length for five years respectively

Years	WAEC(4-Option Length)				NECO (5-Option Length)			
	Bias				Bias			
	Middle B+C		Edge A+D		Middle B+C+D		Edge A+E	
	F	%	F	%	F	%	F	%
2002	20	40.0	30	60.0	37	61.7	23	38.4
2003	26	52.0	24	48.0	38	63.3	22	36.7
2004	29	58.0	21	42.0	34	56.7	26	43.4
2005	26	52.0	24	48.0	35	58.3	25	41.6
2006	22	44.0	28	56.0	37	61.6	23	38.4
Total	123	49.2	127	50.8	181	60.4	119	39.7

Table 1 showed percentage distribution of correct answers keys in WAEC and NECO Economics items between 2002 and 2006. In 2002, a total of 40% of the correct keys were placed in the middle (B+C) as against 60% for the extreme position (A+D). In 2003, 52% in the middle as against 48% at the edge. In 2004, 58% in the middle, while the extreme options were made the keys in 42% of time. In 2005, 52% in the middle and 48% in the extreme. In 2006, 44% in the middle, while 56% in the first and the last positions. Arising from this, it could be observed that the positioning of the keys in 2002 and 2006 was a clear case of edge bias, while in 2003, 2004, and 2005, test writers were middle bias in placing the keys. The implication of this is that the WAEC four-option MC tests do exhibit middle bias more frequently than edge bias. Similarly, in NECO 5-option length, it can be seen that in 2002, 61.7% in the middle (B+C+D) as against 38.4% for the extreme ones (A+E). In 2003, 63.3% in the middle as against 36.7% in the edges. In 2004, 56.7% in the middle, while the extreme options were made the keys in 43.4% of time. In 2005, 58.3% in the middle, 41.6% in the edges. Lastly in 2006, 61.6% in the middle, while 38.4% for first and the last positions. The percentage of time the first and last position was used as keys tally with that of the 2002. Hence, positioning of the correct answers in 5-option NECO Economics MC test items favoured the three middle positions than the extreme positions. It can therefore be concluded the NECO test developers are also middle biased in the positioning of the keys.

Research Hypothesis 1: This hypothesis states that there is no significant influence of primacy and recency positions on the item difficulty of the MC tests.

In testing this hypothesis item analysis was carried out on each of the items in the 2005 adopted and adapted WAEC 50 Economics MC test, after which the difficulty index was grouped into three, Very difficult ($0 \leq P \leq 0.25$), Moderately difficult ($0.26 \leq P \leq 0.74$) and very easy ($0.75 \leq P \leq 1$) according to Hopkins (1998) grouping.

Table 2: Chi-square analysis showing the effects of primacy and recency on item difficulty of 4-option multiple-choice test.

Primacy & Recency	Item Difficulty						Total	χ^2	p
	Very Difficult $0 \leq P \leq 0.25$		Moderately Difficult $(0.26 \leq P \leq 0.74)$		Very Easy $(0.75 \leq P \leq 1)$				
	O	E	O	E	O	E			
A	535	546.76	3162	3249.54	236	136.69	3933	164.16*	<.05
D	469	457.24	2805	2717.45	15	114.31	3289		
Total	1004	1004.00	5967	5966.99	251	251.00	7222		

* Significant at 0.05, O - Observed Value, E - Expected Value

From Table 2, the items whose difficulty level falls within $0 \leq P \leq 0.25$ (very difficult) had options A and D which are the primacy and recency picked 535 and 469 times respectively as the key by the candidates. Also items with moderate difficulty ($0.26 \leq P \leq 0.74$) levels had option A picked as the key 3162 times, while option D was picked 2805 times as the key. For item with low difficulty ($0.75 \leq P \leq 1$), options A and D were picked 236, and 15 times as the key respectively. The chi-square analysis yielded ($\chi^2 = 164.16$) which is significant at the 0.05 level. This implies that primacy and recency positions have significant influence on item difficulty of 4-multiple-choice test. Hence, primacy has more patronage than recency positions in four option MC test.

Table 3: Chi-square analysis showing the effects of primacy and recency options on item difficulty of 5-option multiple-choice test.

Primacy & Recency	Item Difficulty						Total	χ ²	p
	Very Difficult 0 ≤ P ≤ 0.25		Moderately Difficult (0.26 ≤ P ≤ 0.74)		Very Easy (0.75 ≤ P ≤ 1)				
	O	E	O	E	O	E			
A	500	570.81	2789	2714.22	26	29.97	3315	39.69*	<.05
E	300	229.19	1015	1089.78	16	12.03	1331		
Total	800	800.00	3804	3804.00	42	42.00	4646		

From Table 3, primacy options (A), and recency option (E) were picked 500, and 300 times respectively as key very difficult items, while for moderately difficult items, primacy and recency options were picked 2789 and 1015 times respectively. Furthermore in the very easy difficulty level items options A and E were picked as the keys 26 and 16 times respectively. A chi-square analysis of the data yielded $\chi^2 = 39.69$ which is significant at $p < .05$. That is, there is a significant influence of primacy and recency positions on item difficulty in five-option multiple-choice tests. Hence, students patronize primacy than recency in 5-options of the MC test.

Research Hypothesis 2: The second hypothesis states that there is no significant influence of middle attraction on the item difficulty of the multiple-choice test items. Having grouped the items into very difficult, moderately difficult and very easy levels of difficulty, the number of times the students picked the middle options (i.e. B and C) as the key was sorted out. Chi-square analysis was then used to determine the influence of middle attraction on the difficulty level of the items.

Table 4: Chi-square analysis showing the effects of middle attraction on the item difficulty of 4-option multiple-choice test.

Middle Attraction	Item Difficulty						Total	χ^2	p
	Very Difficult $0 \leq P \leq 0.25$		Moderately Difficult $(0.26 \leq P \leq 0.74)$		Very Easy $(0.75 \leq P \leq 1)$				
	O	E	O	E	O	E			
B	597	567.14	3414	3443.85	28	28.00	4039	3.66	>.05
C	537	566.86	3472	3442.14	28	27.99	4037		
Total	1134	1134.00	6886	6886.00	56	55.99	8076		

From Table 4, the chi-square result yielded a value of $\chi^2 = 3.66$ which is not significant at $p > .05$. That is, for 4-option multiple-choice test, the candidates' preference for options in the middle position has no significant influence on the difficulty of the test items. Hence, students patronize middle positions equally in 4-option MC test.

Table 5: Chi-square analysis showing the effects of middle attraction on the item difficulty of 5-option multiple-choice test.

Middle Attraction	Item Difficulty						Total	χ^2	p
	Very Difficult $0 \leq P \leq 0.25$		Moderately Difficult $(0.26 \leq P \leq 0.74)$		Very Easy $(0.75 \leq P \leq 1)$				
	O	E	O	E	O	E	O		
B	604	572.65	3137	3089.78	16	94.56	3757	519.50*	<.05
C	515	585.91	3316	3161.34	13	96.75	3844		
D	504	464.43	2304	2505.88	239	76.69	3047		
Total	1623	1622.99	8757	8757.00	268	268.00	10648		

From Table 5, the chi-square result yielded a values of $\chi^2 = 519.50$ which is significant at $p > .05$. That is, for five-option multiple-choice test, the candidates' preference for options in the middle position has significant influence on the item difficulty of the test. Hence, students prefer to choose the middle point (C) than other middle positions.

Research Hypothesis 3: The third hypothesis states that there is no significant effect of middle attraction on the discrimination level of the MC tests. In testing this hypothesis the discrimination index of the items were determined and grouped into excellent ($0.40 \leq D \leq 1$), good ($0.30 \leq D \leq 0.39$), fair ($0.11 \leq D \leq 0.29$) and poor ($0.01 \leq D \leq 0.10$) discrimination levels in accordance to Hopkins (1998) grouping.

Table 6: Chi-square analysis showing the effects of middle attraction, on item discrimination of 4-option multiple-choice test.

Middle Attraction	Item Discrimination								Total	χ²	p
	Excellent Discrimination (0.40≤D ≤ 1)		Good Discrimination (0.30 ≤ D ≤ 0.39)		Fair Discrimination (0.11 ≤ D ≤ 0.29)		Poor Discrimination (0.01 ≤ D ≤ 0.10)				
	O	E	O	E	O	E	O	E			
B	1866	1841.96	890	988.24	558	567.64	725	642.16	4039	41.93*	<.05
C	1815	1840.04	1086	987.76	577	567.36	559	641.84	4037		
Total	3681	3681.00	1976	1976.00	1135	1135.00	1284	1284.00	8076		

From Table 6, It could be observed that for the excellently discriminating items the two middle option were chosen as the key in a total of 3681 times, while the good discriminating items attracted options B and C as the correct answer in a total of 1976 times. The fair and poorly discriminating items attracted options B and C as the keys in 1125 and 1284 times respectively. The result of the analysis yielded a significant chi-square value ($\chi^2 = 41.93$, $p < .05$). This implies that there is significant effect of middle attraction on item discrimination levels of 4-option item length.

Table 7: Chi-square analysis showing the effect of middle attraction on item discrimination of 5- option item length.

Middle Attraction	Item Discrimination								Total	χ^2	p
	Excellent Discrimination ($0.40 \leq D \leq 1$)		Good Discrimination ($0.30 \leq D \leq 0.39$)		Fair Discrimination ($0.11 \leq D \leq 0.29$)		Poor Discrimination ($0.01 \leq D \leq 0.10$)				
	O	E	O	E	O	E	O	E			
B	1911	1902.85	594	703.20	935	875.74	317	275.21	3757	135.63*	<.05
C	2065	1946.91	639	719.49	843	896.02	297	281.59	3844		
D	1417	1543.24	760	570.31	704	710.24	166	223.20	3047		
Total	5393	5393.00	1993	1993.00	2482	2482.00	780	780.00	10648		

From Table 7, it could be observed that options B, C and D were chosen as the key in 1911, 2065 and 1417 times respectively for items with excellent discrimination. Also for the fairly discriminating items and the poorly discriminating items option B was chosen to be the key in 935 and 317 times respectively while option C was chosen in number of 843 and 297 times and option D in 704 and 166 times. Also, the middle options were chosen as the keys in a total of 1993 times for good discriminating items. A chi-square analysis of the result yielded $\chi^2 = 135.63$, which is significant at $p < .05$. Thus, middle attraction has significant relationship with item discrimination levels in 5-option item length.

8. Discussion

From the results of the data analysis, it was found that the nature of item-key positioning for both WAEC and NECO followed the same trend, but middle bias was more pronounced in 5-option MC format than 4-option MC format. This result is in agreement with research findings, that test developers prefer placing the key in the middle position in 5-option MC format than four-option MC format (Attali & Bar-Hillel, 2001; Haladyna & Downing, 1989 and Inbar, 2002. This study also revealed that positional response bias (e.g. middle bias and edge bias middle) has significant influence on item difficulty of the 4-option and 5-option multiple-choice test. This supported the finding by Attali & Bar-Hillel (2003) that people seek answer to multiple-choice question in the middle position. Another finding of this study shows that there is significant relationship between middle attraction and the discrimination level of Multiple-choice tests in the two test formats. This agreed with the finding of Attali & Bar-Hillel (2003) who found that positional response affects the various discrimination values.

9. Conclusion and Recommendations

The study concluded that test developers tended to exhibit middle bias in placing correct keys in 5-option MC format than in 4-option MC and that key balancing was more pronounced in 4-option MC test.

Based on the findings of this study, the following recommendations were made:

1. Both WAEC and NECO should invite stakeholders in examination bodies to seminars where constant review of items can be discussed.
2. Examination bodies like WAEC and NECO should encourage Universities experts in multiple-choice test items development to participate in generating quality tests items that will be free of any form of biasness be it in the stems, key placement and distractors.
3. NECO should try as much as possible to practise key randomization via the use of computer for easy and quick placement.
4. Both WAEC and NECO should encourage and give privilege to teachers in the secondary schools to participate in seminars that can improve the quality development of teacher's made tests as well as the intricacy of constructing good, quality MC tests.

References

- Afolabi, E.R.I. (1992). Effects of test format, self-concept, and anxiety of response changing behaviour. *Journal of Education and Society*, 2 (1), 39-46.
- Aghassi, M.E. (1990). Getting good grades: how to succeed in college. Englewood Cliffs: Prentices-Hall Inc.
- Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement*, 40, 109-128.
- Bar-Hillel, M., & Attali, Y. (2002). Seek whence: answer sequences and their consequences in key-balanced multiple-choice tests. *The American Statistician*, 56, 299-303.,
- Bennett, R. E. (1993). On the meanings of constructed response. In R. E. Bennet & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.
- Ciardi, F. (2001). Test difficulty, validity, and reliability as functions of selected multiple-choice item construction principles. *Educational Psychological Measurement*, 19, 171-179.
- Clark, E. C. (1956). General response patterns to five-choice items. *Journal of Educational Psychology*, 47, 110-117.
- Fagley, N.S. (1997). Positional response bias in multiple-choice tests of learning: its relation to testwiseness and guessing strategy. *Journal of Educational Measurement*, 17, 241-46.
- Gall, C. & Spurtheim, L. (1999). The effect of keyed response sequencing of multiple-choice items on performance and reliability. *Journal of Educational Measurement*, 21: 45-48.
- Gray, E. William, I. & Henry, F. (2002). Uncertain responded on multiple- choice examinations. *Psychological Reports*, 18, 801-6.
- Haladyna, T.M. & Downing, S.M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2 (1), 51-78.
- Inbar, N. (2002). Middle bias in numerical answers to multiple-choice questions. Unpublished Master's Thesis, the Hebrew University, Jerusalem, Israel.
- Jessell, T. & Sullins, B. (1999). The reliability of a multiple-choice examination under various test-taking instructions. *Journal of Educational Measurement*, 5, 307-14.
- Melfessel, N.S. & Sax, G. (1997). Systematic biases in the keying of correct responses on certain standardized tests. *Educational and Psychological Measurement*, 18, 787-790.
- Mentzer, T. L. (1982). Response biases in multiple-choice test item files. *Educational and Psychological Measurement*, 42, 437-448.
- Rogers, H. J. (1999). Guessing in multiple choice tests. In: Masters, G. N. & Keeves, J. P. (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 235-243). Amsterdam: Pergamon.
- Shaw, J.I., Bergen, J.E., Brown, C.A. & Gallagher, M.E. (2000). Centrality preference in choices among similar options. *The Journal of General Psychology*, 127: 157-164.

